

Structure and inhibition of herpesvirus DNA packaging terminase nuclease domain.

Marta Nadal^{1,2}, Philippe Mas³, Alexandre G. Blanco^{1,2}, Carme Arnan^{1,2}, Maria Solà²,
Darren J. Hart³ and Miquel Coll^{1,2,*}

¹ Institute for Research in Biomedicine, Barcelona Science Park, Baldiri Reixac 10, 08028 Barcelona, Spain.

² Institut de Biologia Molecular de Barcelona (CSIC), Barcelona Science Park, Baldiri Reixac 10, 08028 Barcelona, Spain.

³ EMBL Grenoble, BP 181, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France.

*Corresponding author: E-mail: miquel.coll@irbbarcelona.org. Phone: +34 93 4034951.
Fax: +34 93 4034979.

Classification: BIOLOGICAL SCIENCES, Biochemistry

Author contributions: MN, MS, DJH and MC designed research. MN, PM, AGB and CA performed research; MN, MS, DJH and MC analysed data; MN, MS, DJH and MC wrote the paper.

ABSTRACT

During viral replication, herpesviruses package their DNA into the procapsid by means of the terminase protein complex. In human cytomegalovirus (herpesvirus 5), the terminase is composed by subunits UL89 and UL56. UL89 cleaves the long DNA concatemers into unit-length genomes of appropriate length for encapsidation. We used ESPRIT, a high-throughput screening method, to identify a soluble purifiable fragment of UL89 from a library of 18,432 randomly truncated *ul89* gene constructs. The purified protein was crystallized and its three-dimensional structure was solved. This protein corresponds to the key nuclease domain of the terminase and shows an RNase H/integrase-like fold. We demonstrate that UL89-C has the capacity to process the DNA and that this function is dependent on Mn^{2+} ions, two of which are located at the active site pocket. We also show that the nuclease function can be inactivated by raltegravir, a recently approved anti-AIDS drug that targets the HIV integrase.

\body

INTRODUCTION

Human cytomegalovirus (HCMV) is a member of the herpes family of viruses or *Herpesviridae*. This group includes the human pathogens herpes simplex virus type 1 and 2, varicella zoster virus, Epstein-Barr virus, cytomegalovirus, roseolovirus, and Kaposi's sarcoma-associated herpesvirus. Among these, HCMV, which belongs to the *Betaherpesviridae* subfamily, is widespread throughout the human population and causes the most morbidity and mortality. HCMV infection is rarely serious for people with a competent immune system, although it persists in the host cells and propagates to other individuals. In contrast, infection or reactivation of HCMV is a major cause of life-threatening complications in immunocompromised individuals, such as organ transplant recipients and leukemia or AIDS patients, and is the most significant viral cause of birth defects in industrialized countries (1).

The HCMV genome consists of linear double-stranded (ds) DNA of 230 kb with the highest coding-capacity among *Herpesviridae*. HCMV, like all other herpesviruses, replicates its genomic DNA into high molecular mass head-to-tail concatemers. The newly synthesized multi-copy chains of DNA are then excised into unit-length genomes and each genome is packaged singly into one viral procapsid (2). Maturation into unit-length genome molecules involves viral DNA recognition and cleavage at the site-specific *pac* motifs, which are redundant motifs found at both the 5' and 3' genomic termini (2, 3). The dsDNA endonuclease and packaging activities are performed by a protein complex, the terminase, composed by subunits UL56 and UL89 (4, 5). UL56 has been reported to recognize the *pac* motif (6).

After cleaving one end of the DNA, the terminase translocates the viral DNA into the procapsid, deriving energy for this process through ATP hydrolysis. The procapsid is filled with the DNA molecule and the terminase performs a second dsDNA cleavage, thereby concluding the translocation (3). UL89 has predicted ATPase activity and is most probably the molecular motor for helicase-like DNA translocation (7). In addition, UL89 binds and cleaves DNA molecules, and this activity is enhanced when UL56 is present (5). In agreement, other studies have shown that UL89 interacts specifically with the C-terminal part of UL56 (8). UL56 has the capacity to bind linearized DNA but only upon addition of UL89 is the DNA cut into smaller fragments. This observation indicates that these proteins mediate a concerted reaction of DNA recognition and cleavage (5).

Proteins homologous to UL89 are known in all herpesviruses, as are the other terminase subunits, thus indicating that the DNA packaging mechanism is highly conserved (9). The *ul89* ORF includes two exons separated by a 3902-bp intron. It encodes a two-domain 674 amino acid protein with predicted N-terminal ATPase and C-terminal nuclease activity (Fig. S1) (7). Bacteriophages translocate shorter DNA molecules into their capsids by similar packaging systems (9). Their terminases have been intensively studied, in particular those of phages T4 and RB49(10), λ (11), SPP1(12), P22(13) and Sf6(14), and the structure of the nuclease domains of the large terminase subunits gp17 of RB49 (15) and G2P of SPP1(16) and the full-length large terminase subunit gp17 of T4 (15) have been determined. A theoretical model for the structure of the C-terminal domain of UL89 has been proposed recently (17).

Several studies report inhibitors that prevent the formation of new virions through blockage of the termination system (18-22). However, the structural and functional characterization of herpes packaging proteins, which could assist further discovery and development of novel antiviral molecules, has been hindered by the difficulties in expressing enough soluble material for structural analysis. We have overcome these problems by using ESPRIT (23, 24), a combinatorial library method for defining soluble constructs through random gene truncation and expression screening. This approach yielded a single active soluble construct corresponding to the C-terminal nuclease domain of UL89, the structure of which we report herein at a resolution of 2.15 Å. As such, it is the first protein domain structure reported from the herpesvirus terminase.

RESULTS

Expression of UL89 with a library-based construct screen

The two exons of the HCMV *ul89* gene were cloned as a single DNA construct and initially tested for protein expression in several *E. coli* strains and conditions. No protein obtained from these assays was stable enough to withstand purification. Similar results were obtained when the full-length gene was expressed in insect or mammalian cells. Extensive trials with refolding protocols were also unsuccessful. A number of constructs for each putative domain were designed, based on secondary structure prediction, globularity and disorder, but none of them expressed soluble protein in any system or condition assayed.

Subsequently, in order to find soluble domains, we used the combinatorial library method ESPRIT (23, 24), which generates comprehensive libraries of 5' or 3' truncated genetic constructs of the target. Both libraries were synthesized from the *ul89* gene. We then screened 9216 clones for each library form, corresponding to an approximate four-fold oversample of all possible domain boundaries, for expression of soluble protein. The two libraries were arrayed onto the same nitrocellulose membrane, and colonies were screened for putative soluble protein expression in colony format using measurements of *in vivo* biotinylation efficiency of a C-terminal biotin acceptor peptide by fluorescent streptavidin hybridization (23). Although a relatively large number of clones exhibited positive signals in the 3' truncation library, small-scale 4-ml liquid expression trials yielded only marginally soluble uninteresting fragments of less than 20 kDa in size. In contrast, the 5' truncation library yielded several partially soluble, purifiable constructs of similar size (approximately 37 kDa), from which the 48K22 construct was selected as showing the best behavior following scale-up testing (Fig. S1). This construct was only partially soluble (estimated at 5% of total UL89 protein), but was stable through scale-up to 12 liter culture volumes and yielded approximately 1 mg of purifiable monodisperse protein per liter of culture. Other similar-sized constructs identified as partially soluble in small-scale testing did not maintain solubility during subsequent scale-up steps. Subsequent DNA sequencing and mass spectrometry fingerprinting identified construct 48K22 as a C-terminal fragment of UL89 (residues 418 to 674; Fig. S1), hereafter termed UL89-C. This fragment falls inside the predicted C-terminal nuclease domain, encoded in exon 2.

Overall structure of UL89-C

UL89-C displays a wedged shape with dimensions $40 \times 35 \times 46$ Å. A central eight-stranded mixed β -sheet, with parallel and anti-parallel strands, is flanked by helices α on both sides (Fig. 1A). At one side, hydrophobic interactions pack $\alpha 2$ and $\alpha 3$ against the sheet. At the other site, helices $\alpha 1$, $\alpha 4$, $\alpha 5$ and $\alpha 6$ form a bunch that interacts with the β -sheet by hydrophobic interactions from one side of $\alpha 5$ and $\alpha 6$ and by hydrophilic contacts made by the $\alpha 1$ and $\alpha 4$ C-terminal ends. Two 3_{10} helices, $\eta 1$ and $\eta 2$, at loops connecting $\beta 1$ to $\beta 2$ and $\alpha 6$ to $\beta 10$, border one end of the β -sheet. The strand order in the central sheet is 1, 9, 4, 3, 2, 5, 6, and 10 with topology $+4,-1,-1,+3x,+1x,-5x,+6$ (25)(Fig. 1C and Fig. S2). At both lateral edges of the β -sheet, $\beta 1$ and $\beta 10$ form short strands of only three amino acids each. At one end of the β sheet, long loops surround a cleft that typically harbors the active site in proteins sharing this fold. One of these loops flanking the active site cavity folds in a twisted β -hairpin, formed by $\beta 7$ and $\beta 8$ (Fig. 1A).

UL89-C belongs to the RNase H-like superfamily

A search for structurally similar proteins revealed that UL89-C has the characteristic fold of the RNase H-like superfamily of nucleases and polynucleotidyl transferases (26). The closest structural relatives to UL89-C are the recently reported nuclease domains of the large terminase subunits of bacteriophages, RB49 and T4 gp17 (15)(RMSD 2.6 Å for 158 equivalent C^α and 2.7 Å for 159 equivalent C^α , respectively) and SPP1 G2P(16) (RMSD 3.0 Å for 146 equivalent C^α), the Holliday junction resolvase RuvC(27) (RMSD 2.6 Å for 115 equivalent C^α), the HIV-integrase(28, 29) (RMSD 2.6 Å for 78 equivalent C^α) and the ASV-integrase(30) (RMSD 2.9 Å for 85 equivalent C^α). The crystal structures of all these proteins and other members of the

superfamily display the same basic fold but vary in length and show almost no amino acid sequence identity (i.e., 7.7% identity between UL89-C and the closest structural relative, the nuclease domain of RB49 gp17, after structural alignment). The structural homology between these enzymes can be well described from the structural pattern of human RNase H1 (*Hs*-RNase H1) (31), which consists of a five-stranded β sheet surrounded by α helices on both sides. The order and orientation of the strands within the β -sheet is conserved: 3, 2, 1, 4 and 5, one of them being anti-parallel to the other four ($\uparrow\downarrow\uparrow\uparrow\uparrow$). These strands are equivalent to the UL89-C β -strands 4, 3, 2, 5 and 6, respectively, while helices α A, α B and α E *Hs*-RNase H1 correspond to helices α 2, α 3 and α 6 of UL89-C. All these elements are arranged similarly in all proteins of the superfamily, except for α 6 (α E in *Hs*-RNase H1) which runs in the opposite direction in UL89-C, RB49 gp17, SPP1 G2P and RuvC with respect to the other members of the superfamily (Fig. S3). UL89-C (257 aa) is larger than the bacteriophage homologous proteins gp17 (206 aa) (15) and G2P (178 aa) (16). It is also larger and more complex than RNase H, integrase or resolvase nuclease domains, with the central β -sheet composed by 8 strands rather than 5, and further α helices and other secondary structure elements. Other members of the superfamily, like Tn5 transposase (32, 33) and Piwi-Argonaute (34), also have additional structural elements around the basic RNase H fold, although they are quite different from those found in UL89, thus reflecting their diverse functions, substrates and interactions with other proteins.

Active site cleft

The active site is located at one end of the central β -sheet in a cleft formed by conserved residues, four of them acidic (Fig. 1B). In all structures with a RNase-like fold, the active site is located at a topologically equivalent position, at one end of the β -sheet

where two parallel β -strands ($\beta 2$ and $\beta 5$) separate in a fork-like manner. Asp463, Glu534, and Asp651 coordinate two metal cations (see below). Asp463 is located at the C-terminal end of $\beta 2$ while Glu534 is present at the end of $\beta 5$. Asp651 is found at the beginning of $\alpha 6$, the last α -helix in the structure, which lies diagonally to the two β -strands on one of the faces of the central β -sheet (Fig. 1). These three acidic amino acids are fully conserved and confer a strong electronegative character to the active site (Fig. 2 & Fig. S2). A further conserved aspartate residue, Asp650, is located close to the active site cleft (Fig. 1B) but does not interact directly with any of the metal ions. Comparison of the active site of UL89-C with other RNase-like nucleases (Fig. S3) shows that the presence of several acidic residues coordinating metal ion is a signature of the superfamily, in particular the central residue is always an aspartate (Asp463 in UL89-C). The other residues coordinating the metal may vary. For example in bacteriophage SPP1 G2P(16), one of the closest structural relatives to UL89, an aspartate residue (Asp321) occupies a position equivalent to Glu534 in UL89-C and a histidine residue (His400) that of Asp651. However, in T4 and RB49 gp17(16) the residues coordinating the metal ion are identical to those of UL89-C (Fig. S3). Moreover, in these structures an aspartate residue occupies the equivalent position of Asp650 in UL89-C, although two additional acidic residues of the active pocket present in the bacteriophage structures, Asp 406 and Glu401 in RB49 gp17, are not present in UL89-C.

The active site accommodates two cations

In the crystal not soaked with MnCl_2 , one metal ion was clearly identified at the active site. This corresponds to metal B, as defined by Nowotny & Yang (35). In molecule D of this crystal, an electron density peak initially assigned as a water molecule (the

strongest peak of the water list) could also correspond to another metal ion located at a second position with low occupancy. Indeed, an anomalous difference map calculated from diffraction data from a crystal soaked with MnCl_2 showed two peaks at these two positions (Fig. S4). The first one is coordinated by Asp463 and Glu534 and the second by Asp463 and Asp 651 (Fig. 1B). Asp463, Glu534 and Asp651 (and the closest residues Pro464, Ala465, Gly535, Asn536 and Asp650) are fully conserved among human herpesvirus terminases (Fig. S2), thereby suggesting that they are essential for cation coordination and thus for catalysis. Indeed, Mg^{2+} or Mn^{2+} cations are required for the functioning of these enzymes and a two-metal catalysis has been proposed for their enzymatic mechanism (35, 36).

***In vitro* nuclease assays and mutants**

An *in vitro* assay demonstrated that UL89-C has the capacity to degrade linear and circular DNA and this function is strongly activated by Mn^{2+} (Fig. 3A and B). In the presence of this cation, UL89-C converts supercoiled circular plasmid DNA to nicked open circular DNA, subsequently to linear DNA and finally to completely degraded DNA (Fig. S5). Similarly, UL89-C also degrades linear DNA (Fig. 3A and C). Similar behavior was previously described for the UL89 full-length protein (5). The reaction performed in the same conditions but in the presence of Mg^{2+} instead of Mn^{2+} converted only supercoiled circular plasmid to nicked open circular DNA. With Ca^{2+} , the DNA degradation was even less efficient (Fig. 3A and B). In order to verify that the residues of the structurally-inferred active site were truly involved in the nuclease activity of the protein, we designed a set of single and double mutants and tested their activity. The single mutants D463A, D651A and the double mutant D463A/E534A showed only

residual activity (Fig. 3C and D). These results confirmed that UL89-C harbors the nuclease activity critical for the full length protein performance.

Inactivation by raltegravir

The structural similarity between the herpesvirus terminase nuclease domain and the HIV integrase prompted us to test the inhibitory properties of integrase inhibitors on UL89-C. One of these integrase inhibitors, raltegravir (MK0518), was approved by the FDA in 2007 for the treatment of AIDS (37). Raltegravir turned out to be a strong inhibitor of the nuclease activity of UL89-C (Fig. 4). A recent structure of the prototype foamy virus integrase in complex with DNA and the inhibitor shows that raltegravir binds at the active site, directly coordinating the metal ions (38). Presumably, it would bind in a similar way to UL89-C. In contrast to raltegravir, another integrase inhibitor, elvitegravir (GS9137), had no inhibitory effect on UL89-C under similar conditions.

DISCUSSION

A powerful construct screening technique to obtain UL89-C

UL89, like other herpesvirus DNA packaging proteins, is scarcely expressed in a soluble, purifiable form. Even using insect or mammalian eukaryotic expression systems, we and others were unable to purify a soluble form of this protein in sufficient amounts for crystallographic or even limited proteolysis studies. The ESPRIT (23, 24) analysis reported here permitted the oversampling of all possible domain boundaries as hexahistidine tag fusion positions. However, even this approach resulted in a very low number of soluble expression constructs. This finding is indicative of the challenging nature of UL89. The resulting construct encoding the UL89 C-terminal domain expressed protein that was partially soluble, but the purifiable material was

monodisperse and well-behaving through subsequent concentration and crystallization steps. The identification of this otherwise obscure expression-compatible construct is illustrative of the power of this technique to find rare soluble forms of difficult proteins, and this approach appears particularly effective for viral proteins with uncertain domain boundaries (39).

DNA binding

UL89-C cleaves dsDNA *in vitro* (Fig. 3), as reported previously for the full-length protein (5). This domain should bear the structural determinants for DNA binding. An electrostatic surface calculation indicates that a number of positively charged residues, located in different loops, surround the active site cleft (Fig. 2B). From this calculation, the shape of the surface, and superpositions with *B. halodurans* (40) and human RNase H structures in complex with a DNA/RNA hybrid (31) and Tn5 transposase in complex with DNA (33), we manually built a model for dsDNA bound to UL89-C (Fig. S6). In this model, the loops L β 2- β 3 and L β 5- α 3 fit into the major groove of the DNA, while positively charged side chains appear in close proximity to the phosphates. The sugar-phosphate backbone enters the active site but does not get close enough to the metal ion positions. A distortion (bent) from the regular straight B-DNA used in the model would be necessary for the scissile phosphate to reach the metal ions without clashes of the DNA with the active site surrounding loops. In the bacteriophage large terminase structures, some of these loops, i.e. β 2- β 9 and β 5- α 3, are shorter or less protruding, resulting in a slightly less deep active site. However, Smits *et al.* (16) reported that a protruding β -hairpin in the SPP1 G2P structure would clash with the DNA and suggested that the conformation of this loop changes upon DNA binding. This loop

corresponds to loop $\alpha 5$ - $\alpha 6$ in ULC89-C, which was disordered and not visible in our structure, in agreement with its proposed flexibility.

In RNase H, the equivalent loops to UL89 L $\beta 2$ - $\beta 3$ and L $\beta 5$ - $\alpha 3$ contact the minor groove of the RNA/DNA hybrid, instead of the major groove of the dsDNA, as in our DNA-UL89-C docking model. This observation is not contradictory because the conformation of the RNA/DNA hybrid is a mixture between A and B forms, where the minor groove is wider and the bases are accessible. However, the viral dsDNA is most probably in the B-conformation (albeit probably distorted) where the minor groove is too narrow to permit the entrance of loops L $\beta 2$ - $\beta 3$ and L $\beta 5$ - $\alpha 3$. Therefore, interaction with the bases, if present, would be performed through the major groove. Indeed, loop L $\beta 5$ - $\alpha 3$ is longer and more protruding than its RNase H equivalent. The shallow RNA/DNA hybrid minor groove cannot accommodate this loop, but a deeper B-DNA major groove would fit (Fig. S6).

UL89-C within the terminase complex

In the crystal structure, UL89-C shows four protein molecules in the asymmetric unit, A, B, C and D. Molecules A and B interact with each other about a local two-fold axis, as do molecules C and D (Fig. S7). The interaction surface is at the edge of the central β -sheet so that the sheet extends from one protein to its neighbor. Although UL89 dimers have been detected by cross-linking and gel filtration of the full-length protein (8), UL89-C eluted as a monomer in the size-exclusion chromatography. Thus, with the data available, it is unclear whether the dimer observed in the crystal structure has any physiological relevance or whether it is due to crystal packing. Furthermore, phage and herpesvirus terminases are believed to form toroidal structures and assemble as such

against the 12-fold portal protein (5, 10), for which a dimer like that observed in the crystal structure of UL89-C would not fit. It has been demonstrated by cryo-EM that the phage T4 gp17, a homolog of UL89, forms pentamers (15). In the present structure there is no evidence of ring formation and it is likely that the oligomerization determinants for such an arrangement are outside the UL89-C domain.

UL89 interacts with the UL56 subunit of the terminase. On the basis of results from deletion experiments, the amino acids of UL89 proposed to be involved in this interaction span from residues 580 to 600 (8). This segment corresponds to the exposed helix $\alpha 4$ (Fig. 1 & Fig. S2) and is thus suitable for interaction with UL89 partners. The segment includes three residues that are fully conserved among human herpesvirus, namely Lys583, Ala586 and Asn595. This observation suggests a similar interaction scheme within the family. Furthermore, helix $\alpha 4$ has no counterpart in RNase H or integrases, which are enzymes that do not interact with any protein equivalent to UL56.

UL89 as a drug target

Viral DNA encapsidation machinery has no counterpart in the mammalian cell, thus implying that the proteins involved in this process represent promising selective targets for antiviral therapy. Several studies have reported that inhibitors of DNA packaging in herpesviruses specifically target UL89 and UL56, although the binding sites of the proteins have not been elucidated (18-22). Our study demonstrates that the UL89 C-terminal domain of HCMV and the equivalent domains in all herpesviruses bear the essential nuclease function of the terminase for DNA packaging (Fig. 3). We reveal, for the first time, the three-dimensional structure of this domain in detail and describe the essential residues for the nuclease function, which we demonstrate can be inhibited by

raltegravir, a HIV integrase inhibitor approved by the FDA for AIDS treatment in October 2007 (37). This study therefore opens a way for the design of further optimized inhibitors against UL89-C that may be useful for the development of new anti-herpes drugs.

MATERIALS AND METHODS

Identification of the UL89-C soluble construct from a complete 5' and 3' gene truncation library.

The *ul89* gene from the HHV5 *towne* strain comprises two exons; these were amplified, cloned separately and subsequently ligated together. The library was constructed as described (23). Briefly, for the 5' deletion library, the gene was cloned into a pET9a-derived vector out of frame with a TEV-cleavable N-terminal hexahistidine tag (MGHHHHHHHDYDIPTTENLYFQG) and in frame with a short linker and C-terminal biotin acceptor peptide (SNNGSGGGLNDIFEAQKIEWHE). The presence of *Aat*II and *Asc*I sites between the hexahistidine tag encoding DNA and the *ul89* gene permitted unidirectional truncation of the 5' end of the gene using an exonuclease III. Hexahistidine tag fusions of the truncated gene were generated following recircularization of the plasmid with T4 DNA ligase. Following transformation, the plasmid library was harvested from the *E. coli* cloning strain (Omnimax T1; Invitrogen) and used to transform BL21-CodonPlus-RIL (Stratagene). Robotic processing of the library to identify putative soluble expression constructs was done as described (23, 24). Briefly, 18,432 colonies comprising 9,216 for the 5' and 3' deletion libraries were

picked robotically into microtiter plates of TB broth and grown overnight. These were gridded robotically onto nitrocellulose membranes over LB agar to grow dense colony arrays, which were then induced with IPTG. Colonies were lysed *in situ* and hybridized with Alexa488 Streptavidin (Invitrogen). A fluorimager was used to identify colonies expressing biotinylated proteins. The 96 most intense positive clones of each library were isolated from the library and grown as 4-ml liquid expression cultures. Nickel affinity purifications were performed robotically and purified proteins were assessed by SDS-PAGE.

Purification of wild-type and mutant UL89-C proteins.

E. coli Rosetta cells were transformed with plasmid pHAR-UL89C, pHAR-UL89C-D463A, pHAR-UL89C-D651A and pHAR-UL89C-D463A-E534A where the biotin acceptor peptide had been suppressed by an introduction of a stop codon in the natural position. Cells were grown at 37 °C to an OD₆₀₀ of 0.5, protein expression was induced by addition of IPTG to a final concentration of 0.1 mM with further incubation for 72 h at 16 °C. Cells were harvested at 5000g, re-suspended in binding buffer (50 mM Tris pH 8, 200 mM NaCl, 20 mM imidazole and 200 µl of DNase I at 2 mg/ml) and sonicated. Insoluble material was sedimented by centrifugation (20000g, 4 °C, 25 min), and the supernatant was passed through a 0.45-µm filter. Affinity purification was performed with a 5-ml HisTrap HP column (GE Healthcare). The elution was performed with 20 bed volumes of a linear gradient using a buffer comprising 50 mM Tris pH 8, 200 mM NaCl, and 500 mM imidazole. The fractions were analyzed by SDS-PAGE and pooled. A second chromatographic step was carried out on a Mono Q column (GE Healthcare) using binding buffer consisting of 30 mM Tris pH 8, and 50 mM NaCl. Elution was achieved using a linear salt gradient of 50 mM-1 M NaCl. Fractions were analyzed by

SDS-PAGE and those that showed higher purity were pooled, concentrated and subjected to size-exclusion chromatography using a Superdex 75 10/300 GL column (GE Healthcare). The protein eluted at 10.5 ml, corresponding to 31 kDa. Wild-type and mutant proteins were expressed and purified with similar efficiencies.

Crystallization and heavy-atom derivatization

Protein UL89-C was crystallized by mixing 2 μ l of protein solution containing 10 mg/ml of UL89C, 30 mM Tris buffer (pH 8), 50 mM NaCl and 5 mM EDTA with 2 μ l of precipitant solution containing 10 % (w/v) polyethylene glycol 8000, 150 mM calcium acetate hydrate and 100 mM Mes (pH 6), using the sitting drop vapor diffusion method. Crystals were flash-cooled in 12 % polyethylene glycol 400 as cryoprotectant. To prepare Mn^{2+} -derivatised crystals, native crystals were soaked for 1 h in the crystallization solution enriched with 50 mM MnCl_2 . To prepare Hg heavy-atom derivatives, native protein crystals were soaked for 24 h in the crystallization solution enriched with 0.5 mM ethylmercurithiosalicylic acid sodium salt.

Structure solution and refinement

A native data set was collected at the ESRF ID14-2 beamline to a resolution of 2.15 Å. Crystals belonged to space group $P2_12_12_1$ with cell dimensions $a=82.8$, $b=87.9$, $c=189.4$ and $\alpha=\beta=\gamma=90^\circ$. A data set from a native crystal soaked with Mn^{2+} was collected at ESRF ID29; the crystals belonged to the same space group with the similar cell dimensions. In addition, data for an Hg-derivative were collected at ESRF BM16, at a wavelength of 1.00726 Å (Hg-bound absorption edge). Native and derivative diffraction data were processed using XDS (41), and then scaled, reduced and merged with XSCALE (41) (Table S1). Phases were obtained by Single Isomorphous Replacement

Anomalous Scattering (SIRAS). SHARP (42) was used to determine the positions of 5 Hg atoms using data to 3.5 Å, and phasing the data to 2.15 Å. The resulting map was of insufficient quality for automatic tracing and most of the polypeptide chain had to be built manually using Coot (43). The crystals contained four UL89-C molecules per asymmetric unit. Atomic positions and their associated B-factors were refined with Refmac5 (44) using non-crystallographic symmetry (NCS) restraints. The model was improved by alternating cycles of automatic refinement and interactive model building (Fig. S8). The final refinement cycles included TLS refinement. The Mn²⁺-soaked crystal structure showed two strong electron density peaks at the active site corresponding to the metal ions (Fig. S4). These ions were included and the structure refined with Refmac5. The quality of the stereochemistry of the two structures was assessed with Procheck (45) (Table S1).

***In vitro* nuclease assay**

Purified wild-type and mutant UL89-C domains (final concentration 2 µM) were incubated with 200 ng of circular and linear (digested with HindIII) pUC18 plasmid (2686 bp) in a reaction containing 30 mM Tris pH 8 and 50 mM NaCl for 1 h at 37 °C. The effect of several metal ions was studied by adding 3 mM (final concentration) MgCl₂, CaCl₂ or MnCl₂. The activity was terminated by adding EDTA to a final concentration of 30 mM. The samples were analyzed by agarose gel electrophoresis with ethidium bromide staining. For the inhibitory assay, a range of concentrations of raltegravir (Chemietek) were added to the reaction. A stock solution of 5 mM raltegravir was prepared at 50% DMSO and was further diluted with 30 mM Tris pH 8, 50 mM NaCl to obtain the final concentration.

Miscellaneous

Search for folding relatives was performed with MATRAS (46). Structural alignments and RMSD calculations were performed with SSM (47). Figures 1, S3, S4, S6, S7 and S8 were drawn with Pymol (48). Figure 2 was generate with GRASP (49) and Pymol.

Coordinates. The atomic coordinates for the UL89-C protein structures have been deposited with the Protein Data Base under accession codes 3N4P and 3N4Q.

ACKNOWLEDGEMENTS

We thank Ms. Jenny Colom for help during protein purification, Dr. Jordi Bernués for help during nuclease assays, Ms. Zuzanna Kaczmarek for elvitegravir inhibition assays and Prof. David Stuart for hosting MN to perform expression tests in eukaryotic systems. This study was supported by the *Ministerio de Ciencia e Innovación* (Grant BFU2008-02372/BMC to MC), the *Generalitat de Catalunya* (Grant 2009 SGR 1309 to MC) and the European Commission (Spine2-Complexes LSHG-CT-2006-031220). Synchrotron data collection was supported by the ESRF and the EU. Crystallization screening and preliminary X-ray analysis were performed at the Automated Crystallography Platform, Barcelona.

FIGURE LEGENDS

Figure 1. (A) Overall structure of UL89-C in a ribbon representation. The metal ions are indicated by yellow spheres. UL56-interacting helix $\alpha 4$ is highlighted in blue, α helices in cyan, 3_{10} helices in green and β strands in magenta. (B) Detailed view of the active site area. One Mn^{2+} ion is shown coordinated by Asp463 and Glu534 and a second one is coordinated by Asp463 and Glu651. (C) Topology diagram of UL89. Color code as in (A).

Figure 2. UL89-C surface representations. (A) Conservation surface of UL89-C based on a Risler matrix calculation. Residue color is shown on the basis of conservation score within the eight human herpesvirus: orange, more than 80 % similarity; green, full conservation. Metal ions are shown in yellow. (B) Electrostatic potential surface of UL89-C in the same view as A. The positive potential is shown in blue whereas the negative potential is in red.

Figure 3. UL89-C *in vitro* nuclease assay. (A) Effect of divalent cations on UL89-C linear dsDNA nuclease activity. Lane 1: linear (digested with HindIII) pUC18 plasmid in the absence of the nuclease. Lane 2: nuclease reaction in the absence of divalent cations. Lane 3: nuclease reaction in the presence of Mg^{2+} . Lane 4: nuclease reaction in the presence of Ca^{2+} . Lane 5: nuclease reaction in the presence of Mn^{2+} . (B) Effect of divalent cations on UL89-C circular dsDNA nuclease activity. Lane 1: circular pUC18 plasmid in the absence of the nuclease. Lane 2: linear (digested with HindIII) pUC18 plasmid in the absence of the nuclease. Lane 3: nuclease reaction in the absence of

divalent cations. Lane 4: nuclease reaction in the presence of Mg^{2+} . Lane 5: nuclease reaction in the presence of Ca^{2+} . Lane 6: nuclease reaction in the presence of Mn^{2+} . **(C)** Nuclease assays with the UL89-C WT and mutant proteins with linear dsDNA. Lane 1: linear (digested with HindIII) pUC18 plasmid in absence of the nuclease. Lane 2: nuclease assay with the WT proteins. Lane 3: nuclease assay with the D463A mutant. Lane 4: nuclease assay with the D651A mutant. Lane 5: nuclease assay with the D463A E534A mutant. **(D)** Nuclease assay with the UL89-C WT and mutant proteins with ds-circular DNA. Lane 1: circular pUC18 plasmid in absence of the nuclease. Lane 2: nuclease assay with the WT protein. Lane 3: nuclease assay with the D463A mutant. Lane 4: nuclease assay with the D651A mutant. Lane 5: nuclease assay with the D463A E534A mutant. Mutant proteins show only residual nuclease function. **Labels:** **sc**, supercoiled circular plasmid DNA; **ni-oc**, nicked open circular plasmid DNA; **li**, linear plasmid DNA.

Figure 4. Inhibition of UL89-C nuclease activity by raltegravir. Lane 1: circular pUC18 plasmid in the absence of the nuclease. Lane 2: linear (digested with HindIII) pUC18 plasmid in the absence of the nuclease. Lane 3: nuclease assay with the UL89-C WT protein. Lanes 4-14: nuclease assays in the presence of a range of concentrations of raltegravir (0.5, 1, 2, 4, 8, 16, 32, 64, 128, 256 and 512 μM).

REFERENCES

1. Mocarski ES, Shenk T, & Pass RF (2007) in *Fields Virology*, eds. Knipe DM & Howley PM (Lippincott Williams & Wilkins, Philadelphia), pp. 2701-2772.

2. McVoy MA, Nixon DE, Hur JK, & Adler SP (2000) The ends on herpesvirus DNA replicative concatemers contain pac2 cis cleavage/packaging elements and their formation is controlled by terminal cis sequences. *J. Virol.* 74: 1587-1592.
3. Baines JD & Weller SK (2005) in *Viral genome packaging machines: genetics, structure, and mechanism*, ed. Catalano CE (Springer US, Denver), pp. 135-150.
4. Bogner E (2002) Human cytomegalovirus terminase as a target for antiviral chemotherapy. *Rev. Med. Virol.* 12: 115-127.
5. Scheffczik H, Savva CG, Holzenburg A, Kolesnikova L, & Bogner E (2002) The terminase subunits pUL56 and pUL89 of human cytomegalovirus are DNA-metabolizing proteins with toroidal structure. *Nucl. Acids Res.* 30: 1695-1703.
6. Bogner E, Radsak K, & Stinski MF (1998) The gene product of human cytomegalovirus open reading frame UL56 binds the pac motif and has specific nuclease activity. *J. Virol.* 72: 2259-2264.
7. Champier G, *et al.* (2007) New functional domains of human cytomegalovirus pUL89 predicted by sequence analysis and three-dimensional modelling of the catalytic site DEXDc. *Antivir. Ther.* 12: 217-232.
8. Thoma C, *et al.* (2006) Identification of the interaction domain of the small terminase subunit pUL89 with the large subunit pUL56 of human cytomegalovirus. *Biochemistry* 45: 8855-8863.
9. Przech AJ, Yu D, & Weller SK (2003) Point mutations in exon I of the herpes simplex virus putative terminase subunit, UL15, indicate that the most conserved residues are essential for cleavage and packaging. *J. Virol.* 77: 9613-9621.
10. Sun S, Kondabagil K, Gentz PM, Rossmann MG, & Rao VB (2007) The structure of the ATPase that powers DNA packaging into bacteriophage T4 procapsids. *Mol. Cell* 25: 943-949.
11. Ortega ME, Gaussier H, & Catalano CE (2007) The DNA maturation domain of gpA, the DNA packaging motor protein of bacteriophage lambda, contains an ATPase site associated with endonuclease activity. *J. Mol. Biol.* 373: 851-865.
12. Camacho AG, Gual A, Lurz R, Tavares P, & Alonso JC (2003) Bacillus subtilis bacteriophage SPP1 DNA packaging motor requires terminase and portal proteins. *J. Biol. Chem.* 278: 23251-23259.
13. Nemecek D, *et al.* (2007) Subunit conformations and assembly states of a DNA-translocating motor: The terminase of bacteriophage P22. *J. Mol. Biol.* 374: 817-836.
14. Zhao H, *et al.* (2010) Crystal structure of the DNA-recognition component of the bacterial virus Sf6 genome-packaging machine. *Proc. Natl. Acad. Sci. USA* 107: 1971-1976.
15. Sun S, *et al.* (2008) The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell* 135: 1251-1262.
16. Smits C, *et al.* (2009) Structural basis for the nuclease activity of a bacteriophage large terminase. *EMBO Rep.* 10: 592-598.
17. Couvreur A, *et al.* (2010) Insight into the structure of the pUL89 C-terminal domain of the human cytomegalovirus terminase complex. *Proteins* 78: 1520-1530.
18. Underwood MR, *et al.* (1998) Inhibition of human cytomegalovirus DNA maturation by a benzimidazole ribonucleoside is mediated through the UL89 gene product. *J. Virol.* 72: 717-725.

19. Buerger I, *et al.* (2001) A novel nonnucleoside inhibitor specifically targets cytomegalovirus DNA maturation via the UL89 and UL56 gene products. *J. Virol.* 75: 9077-9086.
20. North TW, Sequar G, Townsend LB, Drach JC, & Barry PA (2004) Rhesus cytomegalovirus is similar to human cytomegalovirus in susceptibility to benzimidazole nucleosides. *Antimicrob. Agents Chemother.* 48: 2760-2765.
21. Dittmer A, Drach JC, Townsend LB, Fischer A, & Bogner E (2005) Interaction of the putative human cytomegalovirus portal protein pUL104 with the large terminase subunit pUL56 and its inhibition by benzimidazole-d-ribonucleosides. *J. Virol.* 79: 14660-14667.
22. Hwang JS, *et al.* (2007) Identification of acetylated, tetrahalogenated benzimidazole d-ribonucleosides with enhanced activity against human cytomegalovirus. *J. Virol.* 81: 11604-11611.
23. Tarendeau F, *et al.* (2007) Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat. Struct. Mol. Biol.* 14: 229-233.
24. Yumerefendi H, Tarendeau F, Mas PJ, & Hart DJ (2010) ESPRIT: An automated, library-based method for mapping and soluble expression of protein domains from challenging targets. *J. Struct. Biol.* In Press, Corrected Proof.
25. Richardson JS (1976) Handedness of crossover connections in beta sheets. *Proc Natl Acad Sci U S A* 73: 2619-2623.
26. Yang W & Sreitz T (1995) Recombining the structures of HIV integrase, RuvC and RNase H. *Structure* 15: 131-134.
27. Ariyoshi M, *et al.* (1994) Atomic structure of the RuvC resolvase: A holliday junction-specific endonuclease from *E. coli*. *Cell* 78: 1063-1072.
28. Dyda F, *et al.* (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* 266: 1981-1986.
29. Maignan S, Guilloteau JP, Zhou-Liu Q, Clément-Mella C, & Mikol V (1998) Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases. *J. Mol. Biol.* 282: 359-368.
30. Lubkowski J, *et al.* (1998) Structure of the catalytic domain of avian sarcoma virus integrase with a bound HIV-1 integrase-targeted inhibitor. *Proc. Natl. Acad. Sci. USA* 95: 4831-4836.
31. Nowotny M, *et al.* (2007) Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol. Cell* 28: 264-276.
32. Rice P & Mizuuchi K (1995) Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell* 82: 209-220.
33. Davies DR, Goryshin IY, Reznikoff WS, & Rayment I (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science* 289: 77-85.
34. Song JJ, Smith SK, Hannon GJ, & Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305: 1434-1437.
35. Nowotny M & Yang W (2006) Stepwise analyses of metal ions in RNase H catalysis from substrate destabilization to product release. *EMBO J.* 25: 1924-1933.

36. Yang W (2008) An equivalent metal ion in one- and two-metal-ion catalysis. *Nat. Struct. Mol. Biol.* 15: 1228-1231.
37. Summa V, *et al.* (2008) Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J. Med. Chem.* 51: 5843-5855.
38. Hare S, Gupta SS, Valkov E, Engelman A, & Cherepanov P (2010) Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 464: 232-236.
39. Ruigrok RW, Crépin T, Hart DJ, & Cusack S (2010) Towards an atomic resolution understanding of the influenza virus replication machinery. *Curr. Opin. Struct. Biol.* 20: 104-113.
40. Nowotny M, Gaidamakov SA, Crouch RJ, & Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* 121: 1005-1016.
41. Kabsch W (1988) Automatic indexing of rotation diffraction patterns. *J. Appl. Cryst.* 21: 67-72.
42. Vonrhein C, Blanc E, Roversi P, Bricogne G, (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.* 364: 215-230.
43. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60: 2126-2132.
44. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53: 240-255.
45. Laskowski RA, MacArthur MW, Moss DS, & Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26: 283-291.
46. Kawabata T (2003) MATRAS: a program for protein 3D structure comparison. *Nucl. Acids Res.* 31: 3367-3369.
47. Krissinel E & Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* 60: 2256-2268.
48. DeLano, W.L. The PyMOL Molecular Graphics System. (2008) DeLano Scientific LLC, Palo Alto, CA, USA. <http://www.pymol.org>.
49. Nicholls A, Sharp KA, & Honig B (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11: 281-296.